



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2019

Orthologous nuclear markers and new transcriptomes that broadly cover the phylogenetic diversity of Acanthaceae

Morais, Erica B ; Schönenberger, Jürg ; Conti, Elena ; Antonelli, Alexandre ; Szövényi, Péter

Abstract: Premise Information on orthologous groups of genes, their sequence variability, and annotation is required for project design in phylogenetic reconstruction. This resource is unavailable for the flowering plant family Acanthaceae (>4000 species). **Methods** We compared transcriptome sequences spanning the extant diversity of Acanthaceae in order to provide a set of orthologous low-copy nuclear genes and assess their utility for reconstructing phylogenetic relationships within this group of plants. **Results** We present new transcriptome assemblies for eight species representing all major clades of Acanthaceae. The assemblies of five of these species are entirely based on new sequence data. Of these five species, three are from subfamilies for which no genomic resources were previously available (Nelsonioideae and Thunbergioideae). These five new transcriptomes are more complete than all others from public databases. Furthermore, we provide alignments with sequence information, annotation, and statistics for potential phylogenetic utility of 1619 orthologous low-copy nuclear markers. **Discussion** Our method of inferring assemblies from multiple pooled tissue samples delivers more complete transcriptomes than any available ones from Acanthaceae. We make available to the community new resources (e.g., sequence information, variability, and annotation of orthologous low-copy nuclear genes) that will help phylogenetic reconstruction in Acanthaceae.

DOI: <https://doi.org/10.1002/aps3.11290>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-174983>

Journal Article

Published Version



The following work is licensed under a Creative Commons: Attribution 4.0 International (CC BY 4.0) License.

Originally published at:

Morais, Erica B; Schönenberger, Jürg; Conti, Elena; Antonelli, Alexandre; Szövényi, Péter (2019). Orthologous nuclear markers and new transcriptomes that broadly cover the phylogenetic diversity of Acanthaceae. *Applications in Plant Sciences*, 7(9):e11290.

DOI: <https://doi.org/10.1002/aps3.11290>

Orthologous nuclear markers and new transcriptomes that broadly cover the phylogenetic diversity of Acanthaceae

Erica B. Morais^{1,5} , Jürg Schönenberger² , Elena Conti¹ , Alexandre Antonelli^{3,4} , and Péter Szövényi¹ 

Manuscript received 10 April 2019; revision accepted 2 August 2019.

¹Department of Systematic and Evolutionary Botany, University of Zurich, 8008 Zurich, Switzerland

²Department of Botany and Biodiversity Research, University of Vienna, Rennweg 14, A-1030 Vienna, Austria

³Gothenburg Global Biodiversity Centre, Box 461, SE 40530 Göteborg, Sweden

⁴Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3AE, United Kingdom

⁵Author for correspondence: ericabarroso@gmail.com

Citation: Morais, E. B., J. Schönenberger, E. Conti, A. Antonelli, and P. Szövényi. 2019. Orthologous nuclear markers and new transcriptomes that broadly cover the phylogenetic diversity of Acanthaceae. *Applications in Plant Sciences* 7(9): e11290.

doi:10.1002/aps3.11290

PREMISE: Information on orthologous groups of genes, their sequence variability, and annotation is required for project design in phylogenetic reconstruction. This resource is unavailable for the flowering plant family Acanthaceae (>4000 species).

METHODS: We compared transcriptome sequences spanning the extant diversity of Acanthaceae in order to provide a set of orthologous low-copy nuclear genes and assess their utility for reconstructing phylogenetic relationships within this group of plants.

RESULTS: We present new transcriptome assemblies for eight species representing all major clades of Acanthaceae. The assemblies of five of these species are entirely based on new sequence data. Of these five species, three are from subfamilies for which no genomic resources were previously available (Nelsonioideae and Thunbergioideae). These five new transcriptomes are more complete than all others from public databases. Furthermore, we provide alignments with sequence information, annotation, and statistics for potential phylogenetic utility of 1619 orthologous low-copy nuclear markers.

DISCUSSION: Our method of inferring assemblies from multiple pooled tissue samples delivers more complete transcriptomes than any available ones from Acanthaceae. We make available to the community new resources (e.g., sequence information, variability, and annotation of orthologous low-copy nuclear genes) that will help phylogenetic reconstruction in Acanthaceae.

KEY WORDS Acanthaceae; Lamiales; nuclear markers; phylogeny; transcriptomes.

Acanthaceae belongs to the order Lamiales (>23,000 species [Chase et al., 2016]). It is among the 15 most species-rich families of flowering plants (>4000 species) and is an ecologically and economically important clade, especially in the tropics (Tripp and McDade, 2014). Despite earlier efforts (e.g., Borg et al., 2008; McDade et al., 2008; Tripp and McDade, 2014), phylogenetic relationships within Acanthaceae remain contentious. In particular, the relationships among major evolutionary lineages within the family (e.g., the four subfamilies) as well as the generic limits within the largest subfamily (Acanthoideae) remain unresolved (e.g., McDade et al., 2018). Furthermore, the phylogenetic relationships in several smaller clades are currently not well understood (e.g., in *Avicennia* L. [Glaser et al., 2019], *Barleria* L. [Darbyshire et al., 2019], *Dyschoriste* Nees [Chumchim et al., 2015], Thunbergioideae T. Anderson [Borg et al., 2008], *Ruellia* L. [Tripp et al., 2018]). In addition to this obstacle, the rooting of the Acanthaceae phylogeny is uncertain, as interfamilial relationships in Lamiales remain elusive (Schäferhoff et al., 2010; Refulio-Rodriguez and Olmstead, 2014; Stull et al., 2015; Wikström et al., 2015; Chase et al., 2016; Sarzi

et al., 2019; Xu et al., 2019). These technical barriers prevent targeted investigations of evolutionary questions within this family.

Plant phylogenies are generally difficult to resolve. This may be due to several reasons, such as low number of substitutions observed in orthologous molecular markers (Li et al., 2019) and/or complex evolutionary histories (e.g., hybridization, polyploidization, complex gene history) not properly dealt with by the algorithms available for phylogenetic inference (e.g., most genes do not fit to any substitution model currently available; Sebastian Höhna, Ludwig-Maximilians Universität, Munich, unpublished data). In addition, gene trees do not necessarily reflect species trees, which further complicates the inference of well-resolved species phylogenies (e.g., Pease et al., 2016). Therefore, to build robust phylogenetic hypotheses, it is essential to compare the phylogenetic inference of several different genes. However, the lack of genomic resources (such as genomic and transcriptomic sequences) in Acanthaceae prevents sequencing of specific target loci. For instance, several molecular markers widely used for plant phylogenetics cannot be amplified in Acanthaceae species

(A. J. Borg and J. Schönenberger, unpublished data). The methods currently accessible for sequencing molecular markers without prior sequence information of orthologous loci are either expensive (e.g., whole genome sequencing) or inappropriate (e.g., restriction site-associated DNA, which has low confidence for homology assessment [i.e., potential paralogy, high levels of missing data, and low reproducibility]) for phylogenetic inference at deeper phylogenetic levels or in older clades. The establishment of low-copy nuclear genes (LCNG) suitable for phylogenetic analysis would help to further clarify the evolutionary history of Acanthaceae and of the Lamiales.

Compared to other currently used strategies of genome reduction prior to sequencing in plant systematics, high-throughput targeted capture (Gnirke et al., 2009) offers several advantages (recently reviewed by Johnson et al., 2019) and has been widely applied in plant systematics and evolution. Transcriptome sequences have been successfully used to develop probe sets for targeting nuclear markers in several plant groups (Chamala et al., 2015; Landis et al., 2016; Crowl et al., 2017; García et al., 2017; Villaverde et al., 2018; Johnson et al., 2019; Vargas et al., 2019). The hybridization between RNA probes and DNA sequences is directly linked to their similarity. Hybridization leads to efficient target enrichment if sequence similarity between RNA probes and DNA sequences shows at least 85% similarity (Orin McCormick, RAPiD Genomics, unpublished data). Therefore, we decided to obtain sequence information for designing specific probes for Acanthaceae.

Currently, there is only a single draft genome published for Acanthaceae (*Ruellia speciosa* Mart. ex Nees, subfamily Acanthoideae; Zhuang and Tripp, 2017), but no genomic resources are available for the subfamilies Thunbergioideae and Nelsonioideae, which comprise some 180 and 172 species, respectively. To maximize the potential for successful hybridization to probe sequences from a set of phylogenetically diverse species such as the Acanthaceae (approximately 80 million years old; Tripp and McDade, 2014), it is critical to include a phylogenetically broad set of taxa when designing probes. In line with this, we generated new transcriptomic data for five species, representing all major clades of Acanthaceae (Fig. 1). Next, we compared our own sequences with transcriptomic data available for Acanthaceae in public data repositories (NCBI Resource Coordinators, 2016). We

provide information on the utility for phylogenetic inference of orthologous loci within this plant group. This study provides a much-needed set of nuclear markers that will facilitate phylogenetic reconstruction within the family Acanthaceae, as well as in Lamiales.

METHODS

Sampling

In order to sample all the major clades (the three first splits in the phylogeny of crown group Acanthaceae, according to Tripp and McDade, 2014) within Acanthaceae (Fig. 1), we carried out RNA sequencing of five species and added data from three additional species from the National Center for Biotechnology Information Short Read Archive (NCBI SRA). We provide new transcriptomic data for two subfamilies lacking genomic resources: (1) Nelsonioideae (*Elytraria carolinensis* (Walter ex J. F. Gmel.) Pers. and (2) Thunbergioideae (*Mendoncia retusa* Turrill and *Thunbergia erecta* (Benth.) T. Anderson). In addition, we sequenced two species that represent major lineages within the subfamily Acanthoideae, *Pachystachys lutea* Nees and *Aphelandra aurantiaca* (Scheidw.) Lindl. (voucher information given in Appendix 1). In order to further increase our sampling (Fig. 1) and to obtain more information on sequence variability, we also retrieved RNA sequencing data for the species *Acanthus leucostachyus* Wall. ex Nees (representing Acantheae), *Andrographis paniculata* (Burm. f.) Nees (representing Andrographideae), and *Avicennia marina* (Forssk.) Vierh. (representing Avicennioideae) from the NCBI SRA (Appendix 1). A recent phylogenomic study comprising Acanthaceae sensu stricto presents a different topology from the one presented by Tripp and McDade (2014) (Amanda Fisher, California State University, unpublished data). However, our sampling still comprises all major clades in the family according to this new topology.

Sample preparation and sequencing

All tissues for RNA sequencing were freshly collected in botanical gardens (see Appendix 1 for voucher information).

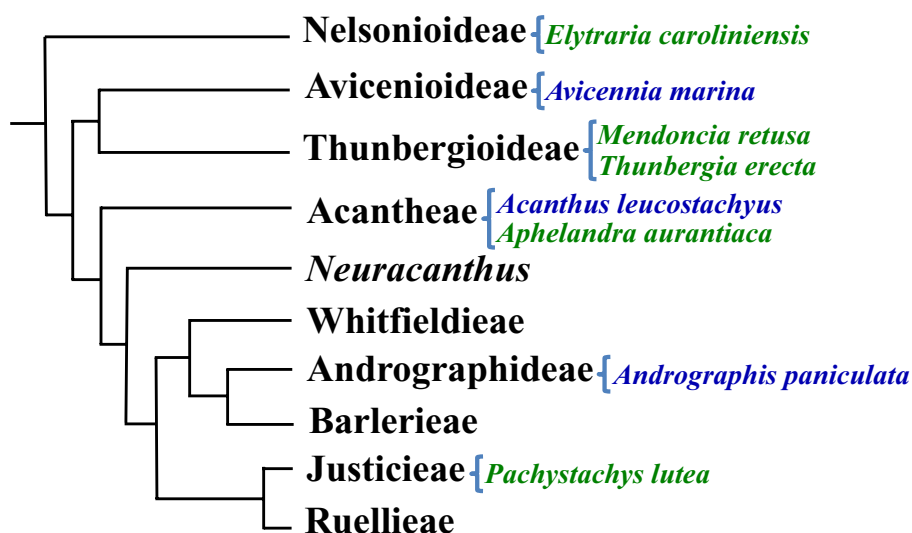


FIGURE 1. Phylogenetic relationships among major clades of Acanthaceae according to Tripp and McDade (2014). Names of species sampled are given on the right side. Taxa for which we provide new data are in green, and taxa for which we used GenBank data are in blue.

Vegetative and reproductive fresh juvenile tissues of each sample were flash-frozen in liquid nitrogen or preserved in RNAlater (Ambion, Waltham, Massachusetts, USA). We attempted to sample as much different young tissues/organs as possible in order to maximize the number of unique genes sequenced. Total RNA was extracted either using the NucleoSpin Plant RNA kit (Macherey-Nagel, Düren, Germany) or the RNeasy Plant Mini Kit (QIAGEN, Hilden, Germany), following the manufacturer's recommendations (Appendix 1).

Illumina TruSeq Stranded mRNA (Illumina, San Diego, California, USA) libraries were prepared (with poly-A RNA enrichment) and sequenced on an Illumina HiSeq2500 sequencer (HiSeq Control Software 2.2.58/RTA 1.18.64) with a 2 × 126-bp setup using HiSeq SBS Kit v4 chemistry (run date 29 January 2016). We pooled all five samples on one lane (output of 249.31 million reads) of one flow cell to generate RNA-Seq reads for each species.

Data cleaning, transcriptome assembly, and annotation

Data quality was visually assessed with FastQC version 0.11.4 (Andrews, 2010) before and after data filtering and trimming. Adapter sequences (the first 13 base pairs), low-quality reads (Phred score < 33), and reads shorter than 50 bp were removed with Trimmomatic-0.35 (Bolger et al., 2014). The sequences were assembled into putative transcripts using Trinity version 2.1.1 (Haas et al., 2013). General statistics for quality assessment of transcriptome assemblies were obtained with the package GenomeTools (Gremme et al., 2013). We used the TRAPID pipeline (Van Bel et al., 2013) based on the PLAZA 2.5 database to get protein translations and to assess the number of fully or quasi-fully sequenced transcripts. We used BUSCO version 3.0.2 (Simão et al., 2015) with the embryophyte single-copy ortholog set to assess the completeness of the transcriptome assemblies. All commands used for this study are available at Figshare (<https://figshare.com/s/7c91497e3fb1cd0ceed7>).

Orthology assessment

To minimize the possibility of obtaining paralogous loci, we aimed at finding orthologous LCNG most appropriate for phylogenomic analyses. These genes are generally highly conserved and are, therefore, not ideal to resolve shallow phylogenetic relationships. However, they often contain introns with greater levels of variability, making them useful for a broad range of phylogenetic analyses even at low taxonomic ranks. We used MarkerMiner (Chamala et al., 2015) to establish groups of orthologous genes by using transcriptome assemblies and their protein translations as input. This

approach uses the same predefined set of genes as a reference and has been successfully applied to capture sequences in other phylogenetic studies in angiosperms (e.g., Nicholls et al., 2015; Landis et al., 2016; Crowl et al., 2017; García et al., 2017; Villaverde et al., 2018; Vargas et al., 2019).

There is currently no taxon closely related to Acanthaceae with a well-annotated high-quality genome available. We were unable to use (with exonerate; Slater and Birney, 2005) the *Ruellia speciosa* genome (Zhuang and Tripp, 2017) as a reference in our analysis due to its low contiguity leading to many fragmentary gene models. Therefore, we decided to use the genome of *Arabidopsis thaliana* (L.) Heynh. as a reference. For each orthogroup recovered by MarkerMiner, we calculated statistics to estimate its phylogenetic utility (e.g., alignment length, number of variable sites, number of parsimony informative sites, AT and GC content) using AMAS (Borowiec, 2016). The output from MarkerMiner gives well-annotated alignments for each orthogroup, including the boundaries of exonic regions in the assembled transcripts (alignments are available at <https://figshare.com/s/9903aacaa3c34bc9ed9>).

RESULTS

Transcriptome assembly

We compared transcriptome assemblies (available at <https://figshare.com/s/aa884dbe565dd1f453b2>) of eight species of Acanthaceae, which represent all major clades (Fig. 1) within this family. Five of these species had no transcriptomic data resources previously available in public databases. Of these five species, three are from subfamilies (Nelsonioideae and Thunbergioideae) for which no genomic resources were available at all. Our transcriptome sequencing resulted in a total of 41–58 million raw reads per sample (Table 1). The assembly of quality-filtered and trimmed reads produced 85,504–286,084 contigs per species. The functional annotation from TRAPID identified 7616–86,113 fully or quasi-fully sequenced transcripts per species (Table 1). Transcriptomes were 83–47.4% complete according to BUSCO (Table 2) (Simão et al., 2015). These five new transcriptomes are more complete than all others from public databases (Table 2).

Phylogenetic utility

We found 1619 putative orthologous LCNGs for Acanthaceae (alignments available at <https://figshare.com/s/9903aacaa3c34bc9ed9>). Here we provide sequence information for bait design in Acanthaceae, offering flexibility of choice based on variability,

TABLE 1. Quantity and quality information of assembled transcriptomes.

Species	Total no. of raw reads	No. of contigs	Mean contig length (bp)	Median contig length (bp)	No. of fully or quasi full-length transcripts	Longest contig (bp)	Shortest contig (bp)
<i>Acanthus leucostachyus</i>	46,888,754	146,742	878.84	470	49,781	12,269	201
<i>Andrographis paniculata</i>	197,537,498	111,881	658.69	426	34,559	6695	201
<i>Aphelandra aurantiaca</i>	47,250,000	127,697	907.29	552	49,509	19,919	201
<i>Avicennia marina</i>	40,000,000	85,504	602.93	395	26,671	6706	201
<i>Elytraria caroliniensis</i>	58,040,000	175,062	959.82	591	7616	15,753	201
<i>Mendoncia retusa</i>	49,650,000	260,725	777.2	457	37,428	16,484	201
<i>Pachystachys lutea</i>	41,080,000	115,380	974.38	584	45,408	14,585	201
<i>Thunbergia erecta</i>	53,290,000	286,084	769.81	450	86,113	16,502	201

TABLE 2. Assessment of transcriptome completeness based on 1440 universal single-copy orthologs (BUSCO).^a

Taxon	Complete	Complete and single-copy	Complete and duplicated	Fragmented	Missing
<i>Acanthus leucostachyus</i> *	78.0%	44.7%	33.3%	6.0%	16.0%
<i>Andrographis paniculata</i> *	59.4%	42.7%	16.7%	16.2%	24.4%
<i>Aphelandra aurantiaca</i>	82.3%	44.7%	37.6%	5.6%	12.1%
<i>Avicennia marina</i> *	47.4%	36.8%	10.6%	20.6%	32.0%
<i>Elytraria caroliniensis</i>	84.3%	28.5%	55.8%	6.0%	9.7%
<i>Mendoncia retusa</i>	82.5%	35.1%	47.4%	6.5%	11.0%
<i>Pachystachys lutea</i>	83.5%	50.0%	33.5%	5.8%	10.7%
<i>Thunbergia erecta</i>	83.0%	34.2%	48.8%	6.6%	10.4%

^aSpecies names with asterisks refer to data downloaded from the National Center for Biotechnology Information; species without an asterisk refer to transcriptomes generated in this study.

presence or absence of species, intron size, and gene size. We make available the set of 1619 alignments from which baits are designed. On average, 3.68 species occurred in each of the orthogroups, which exhibited zero to 0.673 variable sites per position (0.448 on average, median value 0.411). We found 50 orthogroups that included transcripts for all species (eight). This number has increased when orthogroups were required to contain sequence data for at least seven, six, five, and four species (160, 369, 590, and 840, respectively). Parsimony informative sites per orthogroup ranged from zero to 2603 (362 on average, median value 225) (statistics for potential phylogenetic reconstruction available at <https://figshare.com/s/ebb5b55c721debdaccb4>).

DISCUSSION

The exons of all LCNGs (except nine that do not have any variable site: AT1G21370, AT1G31500, AT1G79120, AT3G25530, AT4G01030, AT4G18975, AT4G28830, AT4G38370, and AT5G14140) were found have the potential to solve phylogenetic relationships at deeper nodes (see alignments available at <https://figshare.com/s/9903aacaaa3c34bc9ed9>). We observed that exons are more conserved among species of the same major clade of Acanthaceae (e.g., *Mendoncia* and *Thunbergia*, or *Acanthus* and *Aphelandra*). The variation within Thunbergioideae (*Mendoncia* and *Thunbergia*) is even lower. In order to resolve relationships among closely related species, targeting flanking regions of exons is an efficient approach to sequence the more variable introns (if the DNA is degraded and sequence reads are short, shorter introns are easier to capture and to sequence), which likely provide more phylogenetic information within the Acanthaceae. In this case, longer sequencing reads are desired to sequence introns captured by RNA baits, which are usually designed for exonic regions. Designing RNA primers for PCR is an alternative to hybrid capture. This method is efficient for amplifying long genes, including the introns (e.g., Valderrama et al., 2018).

Vargas et al. (2019) published a python script (GoldFinder) to sub-select markers from the output of MarkerMiner (alignments available at <https://figshare.com/s/9903aacaaa3c34bc9ed9>) according to five relevant criteria for most users who work with molecular phylogenetics and evolution: (1) marker length, (2) percentage of short exons (relative to bait length), (3) number of user’s sequences per marker, (4) similarity, and (5) bait number, length, and coverage. GoldFinder makes the sub-selection task automatic and informed, so that it could be easily run using the data provided on Figshare (<https://figshare.com/s/9903aacaaa3c34bc9ed9>).

Johnson et al. (2019) developed a universal probe set for targeted sequencing of 353 nuclear genes from any angiosperm. However, the efficiency of hybridization of this probe set varied considerably across the different species tested and in some species/clades it was rather low (e.g., 5%, median for all samples was 24.8%). Johnson et al. (2019) selected sequences with up to 30% divergence to design their probes. Accordingly, the capture efficiency for these targets will most likely be improved by designing baits more specific for Acanthaceae, which have not been included in their study. The transcriptome sequences we make available here are a crucial resource for this purpose and can be used to further improve universal probe sets, such as the one by Johnson et al. (2019).

Our method of inferring assemblies from multiple pooled tissue samples delivers more complete transcriptomes than any previously available from Acanthaceae. In addition to being useful for phylogenetic analyses (the main goal of this study), the data generated here provide a potentially important basis for a wide array of other research projects, such as population genomic analyses, metabolic pathway investigations, gene prediction, crop improvement, and analyses of phenotypic diversity. Here, we provide a comparative analysis of representatives of Acanthaceae with the necessary tools for RNA bait design.

DATA AVAILABILITY

The following data are available on Figshare: transcriptome assemblies (<https://figshare.com/s/aa884dbe565dd1f453b2>), alignments for each low-copy nuclear gene (input for baits design) (<https://figshare.com/s/9903aacaaa3c34bc9ed9>), statistics describing their potential for phylogenetic reconstruction (<https://figshare.com/s/ebb5b55c721debdaccb4>), and commands used to perform analyses (<https://figshare.com/s/7c91497e3fb1cd0ceed7>). Raw sequence data are available in the National Center for Biotechnology Information Sequence Read Archive (accession numbers are shown in Appendix 1).

ACKNOWLEDGMENTS

The authors thank the Swiss Confederation (project 2014.0482), the Swiss National Science Foundation (project P1ZHP3_178372 and 160004), the University of Vienna, Claraz Schenkung, Knut and Alice Wallenberg Foundation, the Swedish Research Council, and the Swedish Foundation for Strategic Research for funding. We thank the Science Life Laboratory, the Swedish National Genomics Infrastructure, and Uppmax for assistance in massive parallel

sequencing. We are grateful to the botanical gardens of the University of Gothenburg, Stockholm, and Vienna for plant material.

AUTHOR CONTRIBUTIONS

E.B.M. designed the project, carried out data acquisition and analyses, and wrote the manuscript. J.S. helped to develop the study and supervised the design. P.S. supervised data analyses and writing. All authors have read and revised various versions of the manuscript.

LITERATURE CITED

- Andrews, S. 2010. FastQC: A quality control tool for high throughput sequence data. Website <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> [accessed 11 February 2016].
- Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics* 30(15): 2114–2120.
- Borg, A. J., L. A. McDade, and J. Schönenberger. 2008. Molecular phylogenetics and morphological evolution of Thunbergioideae (Acanthaceae). *Taxon* 57: 811–822.
- Borowiec, M. L. 2016. AMAS: A fast tool for alignment manipulation and computing of summary statistics. *PeerJ* 4: e1660.
- Chamala, S., N. García, G. T. Godden, V. Krishnakumar, I. E. Jordon-Thaden, R. De Smet, W. B. Barbazuk, et al. 2015. MarkerMiner 1.0: A new application for phylogenetic marker development using angiosperm transcriptomes. *Applications in Plant Sciences* 3(4): 1400115.
- Chase, M. W., M. J. M. Christenhusz, M. F. Fay, J. W. Byng, W. S. Judd, D. E. Soltis, D. J. Mabberley, et al. 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society* 181: 1–20.
- Chumchim, N., L. McDade, and A. E. Fisher. 2015. Phylogeny of *Dyschoriste* (Acanthaceae). *Aliso* 33(2): 77–89.
- Crowl, A. A., C. Myers, and N. Cellinese. 2017. Embracing discordance: Phylogenomic analyses provide evidence for allopolyploidy leading to cryptic diversity in a Mediterranean *Campanula* (Campanulaceae) clade. *Evolution* 71(4): 913–922.
- Darbyshire, I., A. E. Fisher, C. A. Kiel, and L. A. McDade. 2019. Phylogenetic relationships among species of *Barleria* (Acanthaceae, Lamiales): Molecular data reveal complex patterns of morphological evolution and support a revised classification. *Taxon* 68(1): 92–111.
- García, N., R. A. Folk, A. W. Meerow, S. Chamala, M. A. Gitzendanner, R. S. Oliveira, D. E. Soltis, and P. S. Soltis. 2017. Deep reticulation and incomplete lineage sorting obscure the diploid phylogeny of rain-lilies and allies (Amaryllidaceae tribe Hippeastreae). *Molecular Phylogenetics and Evolution* 111: 231–247.
- Glasenapp, Y., I. Korth, X.-V. Nguyen, and J. Papenbrock. 2019. Sustainable use of mangroves as sources of valuable medicinal compounds: Species identification, propagation and secondary metabolite composition. *South African Journal of Botany* 121: 317–328.
- Gnirke, A., A. Melnikov, J. Maguire, P. Rogov, E. M. LeProust, W. Brockman, T. Fennell, et al. 2009. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnology* 27: 182–189.
- Gremme, G., S. Steinbiss, and S. Kurtz. 2013. GenomeTools: A comprehensive software library for efficient processing of structured genome annotations. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 10(3): 645–656.
- Haas, B. J., A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M. B. Couger, et al. 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols* 8: 1494–1512.
- Johnson, M., L. Pokorny, S. Dodsworth, L. R. Botigué, R. S. Cowan, A. Devault, W. L. Eiserhardt, et al. 2019. A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering. *Systematic Biology* 68(4): 594–606.
- Landis, J. B., R. D. O'Toole, K. L. Ventura, M. A. Gitzendanner, D. G. Oppenheimer, D. E. Soltis, and P. S. Soltis. 2016. The phenotypic and genetic underpinnings of flower size in Polemoniaceae. *Frontiers in Plant Sciences* 5: 1144.
- Li, H. T., T. S. Yi, L. M. Gao, P. F. Ma, T. Zhang, J. B. Yang, M. Gitzendanner, et al. 2019. Origin of angiosperms and the puzzle of the Jurassic gap. *Nature Plants* 5: 461–470.
- McDade, L. A., T. F. Daniel, and C. A. Kiel. 2008. Toward a comprehensive understanding of phylogenetic relationships among lineages of Acanthaceae s.l. (Lamiales). *American Journal of Botany* 95(9): 1136–1152.
- McDade, L. A., T. F. Daniel, and C. A. Kiel. 2018. The *Tetramerium* lineage (Acanthaceae, Jussiceae) revisited: Phylogenetic relationships reveal polyphyly of many New World genera accompanied by rampant evolution of floral morphology. *Systematic Botany* 43(1): 97–116.
- NCBI Resource Coordinators. 2016. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research* 44 (Database issue): D7–D19.
- Nicholls, J. A., R. T. Pennington, E. J. M. Koenen, C. E. Hughes, J. Hearn, L. Bunnefeld, K. G. Dexter, et al. 2015. Using targeted enrichment of nuclear genes to increase phylogenetic resolution in the neotropical rain forest genus *Inga* (Leguminosae: Mimosoideae). *Frontiers in Plant Sciences* 6: 710.
- Pease, J. B., D. C. Haak, M. W. Hahn, and L. C. Moyle. 2016. Phylogenomics reveals three sources of adaptive variation during a rapid radiation. *PLoS Biology* 14(2): e1002379.
- Refugio-Rodriguez, N. F., and R. G. Olmstead. 2014. Phylogeny of Lamiidae. *American Journal of Botany* 101(2): 287–299.
- Sarzi, D. S., L. Haerolde, F. S. Lopes, C. Furtado, D. R. Oliveira, C. M. Sakuragui, and F. Prosdocimia. 2019. Complete plastid genome of *Lippia organoides* (Verbenaceae) and phylogenomic analysis of Lamiales. *Mitochondrial DNA Part B* 4(1): 808–810.
- Schäferhoff, B., A. Fleischmann, E. Fischer, D. C. Albach, T. Borsch, G. Heubl, and K. F. Müller. 2010. Towards resolving Lamiales relationships: Insights from rapidly evolving chloroplast sequences. *BMC Evolutionary Biology* 10: 352.
- Simão, F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, and E. M. Zdobnov. 2015. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31(19): 3210–3212.
- Slater, G. S., and E. Birney. 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6: 31.
- Stull, G. W., R. D. S. Stefano, D. E. Soltis, and P. S. Soltis. 2015. Resolving basal lamiid phylogeny and the circumscription of Icacinaceae with a plastome-scale data set. *American Journal of Botany* 102(11): 1794–1813.
- Tripp, E. A., and L. A. McDade. 2014. A rich fossil record yields calibrated phylogeny for Acanthaceae (Lamiales) and evidence for marked biases in timing and directionality of intercontinental disjunctions. *Systematic Biology* 63(5): 660–684.
- Tripp, E. A., Y. Zhuang, M. Schreiber, H. Stone, and A. E. Berardi. 2018. Evolutionary and ecological drivers of plant flavonoids across a large latitudinal gradient. *Molecular Phylogenetics and Evolution* 128: 147–161.
- Valderrama, E., J. E. Richardson, C. A. Kidner, S. Madrián, and G. N. Stone. 2018. Transcriptome mining for phylogenetic markers in a recently radiated genus of tropical plants (*Renealmia* L.f., Zingiberaceae). *Molecular Phylogenetics and Evolution* 119: 13–24.
- Van Bel, M., S. Proost, C. Van Neste, D. Deforce, Y. Van de Peer, and K. Vandepoele. 2013. TRAPID, an efficient online tool for the functional and comparative analysis of *de novo* RNA-Seq transcriptomes. *Genome Biology* 14: R134.
- Vargas, O. M., M. Heuertz, S. A. Smith, and C. W. Dick. 2019. Target sequence capture in the Brazil nut family (Lecythidaceae): Marker selection and in silico capture from genome skimming data. *Molecular Phylogenetics and Evolution* 135: 98–104.
- Villaverde, T., L. Pokorny, S. Olsson, M. Rincón-Barrado, M. G. Johnson, E. M. Gardner, N. J. Wickett, et al. 2018. Bridging the micro- and macroevolutionary levels in phylogenomics: Hyb-Seq solves relationships from populations to species and above. *New Phytologist* 220: 636–650.
- Wikström, N., K. Kainulainen, S. G. Razafimandimbison, J. E. E. Smedmark, and B. Bremer. 2015. A revised time tree of the Asterids: Establishing a temporal

- framework for evolutionary studies of the coffee family (Rubiaceae). *PLoS ONE* 10(5): e0126690.
- Xu, W.-Q., J. Losh, C. Chen, P. Li, R.-H. Wang, Y.-P. Zhao, Y.-X. Qiu, and C.-X. Fu. 2019. Comparative genomics of figworts (*Scrophularia*, Scrophulariaceae), with implications for the evolution of *Scrophularia* and Lamiales. *Journal of Systematics and Evolution* 57(1): 55–65.
- Zhuang, Y., and E. A. Tripp. 2017. The draft genome of *Ruellia speciosa* (Beautiful Wild Petunia: Acanthaceae). *DNA Research* 24: 179–192.

APPENDIX 1. Sample details for taxa used in this study.

Taxon	Tissue	Fixation method	Extraction method	Voucher ^a	SRA accession no.	Botanical garden
<i>Acanthus leucostachyus</i> Wall. ex Nees	L, R	—	—	—	SRR1793319*	—
<i>Andrographis paniculata</i> (Burm. f.) Nees	L	—	—	—	SRR1292497*	—
<i>Aphelandra aurantiaca</i> (Scheidw.) Lindl.	L, Fl	RNA ^{later}	QIAGEN	E.B. Morais 168, WU 0104498	SRR8782583	University of Vienna
<i>Avicennia marina</i> (Forssk.) Vierh.	L	—	—	—	SRR653719*	—
<i>Elytraria caroliniensis</i> (Walter ex J. F. Gmel.) Pers.	L, Fl	Liquid nitrogen	QIAGEN	E.B. Morais 163, WU 0104492	SRR8756096	University of Vienna
<i>Mendoncia retusa</i> Turrill	S, L, Fl	Liquid nitrogen	Macherey-Nagel	A.J. Borg 1, S or Z 000172480	SRR8749657	University of Stockholm
<i>Pachystachys lutea</i> Nees	S, L, Fl	RNA ^{later}	QIAGEN	E.B. Morais 164, WU 0104493	SRR8755478	University of Vienna
<i>Thunbergia erecta</i> (Benth.) T. Anderson	S, L, Fl	Liquid nitrogen	Macherey-Nagel	E.B. Morais 173, Z 000172481	SRR8752193	University of Gothenburg

Note: — = not available; Fl = flower; L = leaf; R = root; S = stem.

^aHerbaria abbreviations follow Index Herbariorum (see <http://sweetgum.nybg.org/science/ih/>).

*Data downloaded from the National Center for Biotechnology Information.